



Thermal integrity profiling of concrete piles incorporating three-dimensional defects using deep convolutional neural networks

Javier Sánchez Fernández , Agustín Ruiz López & David M. G. Taborda

To cite this article: Javier Sánchez Fernández , Agustín Ruiz López & David M. G. Taborda (22 Jun 2026): Thermal integrity profiling of concrete piles incorporating three-dimensional defects using deep convolutional neural networks, Geomechanics and Geoengineering, DOI: [10.1080/17486025.2026.2691953](https://doi.org/10.1080/17486025.2026.2691953)

To link to this article: <https://doi.org/10.1080/17486025.2026.2691953>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Jun 2026.



[Submit your article to this journal](#)



Article views: 350



[View related articles](#)



[View Crossmark data](#)

Thermal integrity profiling of concrete piles incorporating three-dimensional defects using deep convolutional neural networks

Javier Sánchez Fernández^{a,b}, Agustín Ruiz López^{b,c} and David M. G. Taborda^b

^aScience and Solutions for a Changing Planet DTP, Imperial College London, London, UK; ^bDepartment of Civil and Environmental Engineering, Imperial College London, London, UK; ^cSeequent, The Bentley Subsurface Company, Delft, The Netherlands

ABSTRACT

Thermal integrity profiling (TIP) has emerged as a powerful non-destructive method for evaluating cast-in-place concrete piles by monitoring heat of hydration (HoH) temperature profiles using temperature probes embedded in the piles. However, conventional TIP interpretation relies either on manual inspection of a series of temperature-depth profiles over time, which is inherently prone to subjective bias, or on back-analysis using simplified modelling approaches. This study presents, for the first time, a fully three-dimensional TIP framework combining finite element (FE) simulations with two deep neural networks (NN). A multi-label classifier detects up to three defects per pile, while a regressor quantifies defect volume, location, reinforcement-cage displacement and hydration parameters. A database with hydration temperature profiles from 19,445 simulations encompassing randomised pile lengths, soil stratigraphies, cage misalignments and multi-defect configurations is used to train the two NNs. The classifier achieves a per-defect accuracy of 93.6% (96.8% per sample). The regression model achieves near-perfect accuracy for cage displacement ($R^2 = 0.992$) and defect volume (mean absolute error of 0.03 m^3), and an adjusted R^2 of 0.92 for defect location. This novel automated approach significantly enhances the objectivity, test speed and reliability of TIP-based defect detection across a wide range of deep pile foundations.

ARTICLE HISTORY

Received 28 January 2026
Accepted 15 June 2026

KEYWORDS

Thermal integrity profiling;
defect detection; machine
learning; concrete piles;
finite element modelling

1. Introduction

Concrete is the most widely used construction material worldwide, with an estimated annual consumption of 30 billion tonnes (Monteiro *et al.* 2017). The production of cement it requires is responsible for about 8% of global carbon emissions or 3 billion tonnes of CO₂ (Wu *et al.* 2024). Therefore, civil engineers have the responsibility to ensure its efficient, durable and safe application, particularly in large-scale infrastructure projects. One of the largest uses of concrete in buildings occurs within their foundation systems, particularly in cast-in-place elements such as bored piles. A survey of 200 buildings by Pratt (2016) shows that foundation materials alone account for approximately 20–25% of a building's total weight, illustrating the substantial proportion of concrete allocated to the substructure. This magnitude reflects both the demanding geotechnical requirements of load transfer and settlement control and the predominance of concrete as the material of choice for deep foundations. Therefore, any improvement in the modelling, optimisation or monitoring of concrete foundation systems has the potential to influence a significant share of the total material use in buildings.

Despite careful design and construction practices, defects in concrete piles can still occur. Common anomalies include necking, bulging, rebar cage misalignments, air voids and soil inclusions (Loseva *et al.* 2022). In a study carried out on 2% of all cast-in-place piles in Germany, 15% of the tested piles showed a deviation from the ideal result, with 5% of the total being labelled as defective (Klingmuller and Kirsch 2004). It is therefore essential to test new bored pile foundations, as these defects can severely compromise load-bearing capacity, long-term durability and structural performance.

However, given the challenges inherent to the inspection of buried structures, the detection of these defects can be difficult. Some common conventional non-destructive testing (NDT) methods include cross-hole sonic logging (CSL) or low-strain integrity testing (ASTM 2016a, 2016b). While these methods provide

a fairly straightforward approach to NDT, they also present some important drawbacks and limitations to be considered. Boeckmann and Loehr (2019) explain the lack of prediction capabilities of CSL outside the reinforcement cage, as well as its tendency to yield a large number of false positives. According to Finno *et al.* (2002), low-strain testing can provide an overall assessment of the pile, but lacks detailed, reliable output.

Thermal Integrity Profiling (TIP) offers a new non-destructive alternative for assessing concrete pile quality during early curing. First introduced by Mullins *et al.* (2007), TIP is based on recording temperature changes during the exothermic hydration process of fresh concrete. Thermal sensors, typically attached to the reinforcement cage, record the temperature evolution over time, and these profiles are analysed to infer potential anomalies. Since the presence of a defect, such as a soil inclusion or air pocket, alters the heat of hydration (HoH) generated in the area, deviations in the expected thermal profile are used to infer the existence of irregularities.

Traditionally, interpreting TIP data requires manual visual inspection of temperature-depth or temperature-time plots. This process is time-consuming, subjective and highly dependent on the experience of the practitioner (Qiu *et al.* 2025). Furthermore, it does not rigorously account for variations in soil thermal conductivity, concrete hydration kinetics or geometric configuration of the pile. As a result, conventional analysis provides only an approximate assessment of potential defects and cannot quantify their location or severity with confidence. A step forward has been made in recent years, where numerical modelling has become central to advancing the interpretation of TIP data beyond visual inspection or empirical correlations. One of the earliest developments was the integration of distributed fibre-optic sensing (DFOS) with one-dimensional axisymmetric finite element heat-transfer models, which enabled back-analysis of curing temperatures to infer variations in pile diameter and concrete cover while accounting for soil stratigraphy and thermal boundary conditions (Rui *et al.* 2017). This demonstrated that measured thermal profiles could be matched to predictions by varying soil and concrete parameters, providing a more rigorous basis for detecting deviations arising from defects or cage eccentricity. Building on this foundation, Sun *et al.* (2021) introduced a more systematic interpretation framework that explicitly linked TIP measurements with cement hydration through two-dimensional FE simulations, allowing defects to be localised and sized by comparing field measurements with numerically generated profiles. Their approach incorporated calibrated hydration models, consideration of soil layering, cage eccentricity and thermal property variability, all of which are factors known to distort empirical TIP readings but previously unaccounted for in industry practice.

More recently, the use of numerical modelling has expanded into fully three-dimensional FE frameworks capable of representing the spatially variable nature of heat generation and diffusion in cast-in-place piles. Mahjoubi *et al.* (2025) developed an advanced 3D modelling approach in which cement hydration is simulated on an element-by-element basis, enabling accurate simulation of phenomena that cannot be captured in 1D or 2D formulations. Their parametric and statistical analyses further revealed that concrete-specific heat, and concrete and soil thermal conductivities, are the dominant factors governing TIP sensitivity, and that using fixed empirical values (as is common in the current TIP software) can lead to misinterpretation of anomaly size or even false positives. These developments mark a significant shift from qualitative assessment towards physics-based, quantitatively calibrated analysis frameworks capable of delivering much more reliable defect detection and characterisation than traditional TIP practice.

Given the considerable advantages of TIP for the evaluation of concrete pile foundations, there is a need for a more systematic, accurate and automated approach to interpret thermal data and identify, in a fast and rigorous manner, not only the existence of anomalies but also their volume and position within the pile. In the current practice, as demonstrated in field implementations such as those reported by Mullins (2010), interpretation relies on temperature versus depth profiles and peak detection combined with engineering judgement. While effective for identifying pronounced anomalies, this approach does not directly provide quantitative estimates of defect geometry and location.

To address this gap, the present study introduces a data-driven framework for TIP interpretation accounting for three-dimensional effects based on two artificial neural networks. These models are intended to serve as 'general models' that can address a variety of soil conditions, concrete types, pile shapes and defect scenarios. The aim of this work is to showcase the possibilities offered by the chosen approach. In the future, users should tailor these models to better match the specifics of their own tests, which may require

deeper insight into the characteristics of concrete mixes, the thermal properties of soils of interest, and especially how temperature will be monitored across the pile's cross-section. The proposed method operates directly on temperature measurements obtained from standard TIP probe layouts and is therefore compatible with existing field-testing procedures. The first model performs multi-label classification to determine the presence or absence of up to three defects per pile, while the second estimates key defect characteristics, including spatial location, volume and reinforcement cage misalignment, in a fast, consistent and automated manner, extending beyond traditional profile-based interpretation or simplified numerical modelling.

2. Proposed approach

As outlined above, the working principle behind thermal integrity profiling (TIP) is to detect potential defects within concrete pile foundations by identifying unexpected variations in temperature distributions arising from the hydration of concrete. Despite recent advances, existing predictive models have not yet fully captured the combined effect of concrete and soil characteristics, geometric variability and cage positioning within a three-dimensional framework.

Previous approaches to TIP analysis have predominantly relied on visual inspection of temperature profiles along the pile depth to infer the presence of defects, as these were regarded as sufficient (Mullins and Winters 2011). More recently, 2D HoH simulations of concrete piles have been employed to generate large datasets for training machine learning models, e.g. Sánchez Fernández (2025a). These datasets were used to develop a two-stage neural network framework: an artificial neural network (ANN) classifier to detect the presence of defects, followed by a regressor to estimate the associated defect properties (Sánchez Fernández *et al.* 2025). While this marked an improvement in the quantitative interpretation of TIP data, the methodology was constrained by its reliance on two-dimensional cross-sectional representations of the pile.

This two-dimensional simplification introduces several limitations in terms of real-world applicability. Specifically, it assumes that the pile depth and the dimension of the defect in the axial direction have no measurable impact, as they are effectively modelled as infinite in the out-of-plane direction, thereby preventing randomisation of defect location along the pile axis and potentially facilitating the task of identifying the defect. Furthermore, it restricts the simulation to a single homogeneous soil layer and allows, at most, one defect per pile. While the limitations related to soil variability and defect properties were partially mitigated through a wide parametric variation, the inability to model arbitrary defect depths and multi-defect scenarios remains a fundamental shortcoming. These can only be fully addressed through a three-dimensional formulation that captures the spatial complexity and geometric variability inherent to real-world TIP applications.

To overcome these limitations, the present study builds upon and extends the methodology proposed in Sánchez Fernández *et al.* (2025), introducing a refined framework capable of processing three-dimensional TIP data. This approach enables the prediction of several defects in concrete piles that reflect the geometric and material complexities encountered in practice with greater accuracy.

The proposed framework comprises two deep learning models: a classification model to predict the presence or absence of up to three defects per pile and a regression model to estimate detailed defect properties, reinforcement cage displacement and concrete hydration characteristics. Both models are based on deep convolutional neural networks (CNNs). This dual-model architecture is designed to operate effectively across a broad spectrum of pile geometries, stratified soil profiles, concrete compositions and defect scenarios, providing a flexible and general solution which can be scaled for practical integrity assessment of foundation systems comprising a large number of piles.

2.1. Finite element modelling

Several modelling strategies are available in the literature for simulating the HoH in concrete, including models based on cement's chemical reaction kinetics, multi-phase transport formulations, or empirical data fitting (Parrot and Killoh 1984, Swaddiwudhipong *et al.* 2002, Schindler and Folliard 2005). In this study, the heat of hydration model presented in Liu *et al.* (2022) was chosen due to its ability to predict a very wide range of hydration curves using only three parameters, namely τ , β and Q_{max} . These control, respectively,

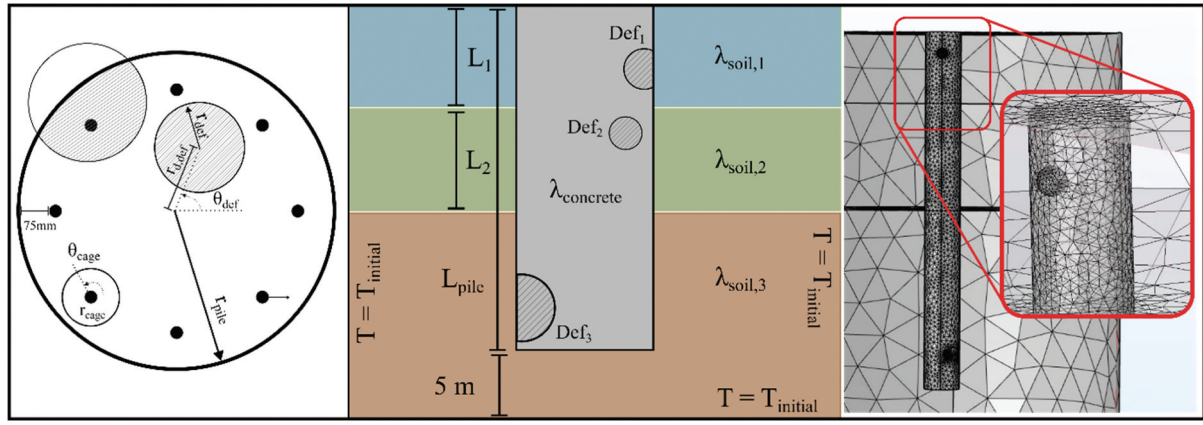


Figure 1. Schematics showing a representative pile cross-section (left), model's geometry, variables and boundary conditions (middle) and typical pile mesh resolution (right).

the rate at which hydration happens, when hydration starts, and the maximum heat production available in the concrete mix. The model implementation was verified by reproducing representative hydration curves, confirming consistency with the original formulation. Additional details on the formulation and validation of this hydration model can be found in Liu *et al.* (2022), while an example of its implementation for practical use in COMSOL Multiphysics is described in Sánchez Fernández *et al.* (2025).

The hydration model was implemented using the heat transfer module in COMSOL Multiphysics and used to produce evolutions of temperatures over time across different material and geometrical configurations in three-dimensional analyses. In order to keep the model parametrisable, and to accommodate for the different defect configurations, the FE model was created using COMSOL's Livelink for MATLAB. This allowed a fully automated generation of geometry, material properties, boundary conditions and meshing parameters for each simulation, as well as integration with the data generation pipeline. Heat of hydration was only generated in areas within the pile without defects.

The modelled geometry corresponded to a cylindrical soil domain of 20 m diameter and depth of $L_{pile} + 5$ m, centred on a vertical reinforced concrete pile. The domain size was chosen to be sufficiently large to prevent any interference of the boundary conditions during the hydration process, and a Dirichlet boundary condition (i.e. constant temperature) was assumed for all external boundaries. The initial model temperature for each simulation was set to the simulation's specific pouring temperature. Defects were represented as spherical voids randomly embedded within the pile, being trimmed by the pile wall face, generating lens-like shapes if these extended beyond the pile domain.

To account for potential stratification and variability in soil thermal response, the surrounding soil domain was modelled with three horizontal layers of randomised thickness and thermal properties. Each of the corresponding three layers within the pile has the possibility of having either one or no defect, therefore modelling between zero and three defects per simulation.

Time-stepping was performed using fixed output steps to avoid temporal interpolation, with a maximum time step of 0.05 days, a relative tolerance of $1e-5$ and a total simulation time of 5 days.

Meshing was carried out using an in-built mesher, with additional refinement zones introduced near defect boundaries and concrete pile to capture the steeper thermal gradients. Second-order tetrahedral elements were used throughout all domains, with an increased resolution in regions where greater thermal gradients were expected, such as the pile itself and the volume of soil immediately adjacent to it. Given the complex geometric arrangements that can arise due to the randomisation of defect location and volume, further refinements were used when the distance between domain boundaries (e.g. defect and pile wall) was limited. Based on these principles adopted for mesh generation, which were established based on comprehensive mesh sensitivity studies, each case was discretised using, on average, approximately 50,000 elements, using a minimum element size of 0.05 m within the pile and in its proximity, increasing to a maximum element size of 2.5 m close to the external soil boundary. A representative pile cross-section, geometry, soil strata, boundary conditions and mesh are shown in Figure 1.

2.2. Numerical database

A dataset comprising 19,445 unique simulations (samples) was generated to support the training and evaluation of the classification and regression models. The samples comprise a range of defect scenarios, including 2449 samples with three defects, 5231 with two, 5373 with one and 6392 with no defect. It is assumed that, at most, only one defect can occur in the pile section embedded into a given soil layer. Therefore, since three soil layers exist, all possible combinations of the presence of defects across the three pile sections were included. For instance, in samples with two defects, configurations such as (0-1-1), (1-0-1) and (1-1-0) were considered, where '1' indicates the presence of a defect in the pile section embedded in a given soil layer and '0' indicates its absence. Note that while three soil layers were modelled, the fact that their properties are randomly allocated during the sampling process means that a quasi-uniform soil deposit is, in effect, possible.

Each simulation produces a spatio-temporal field of temperature increase relative to the initial temperature, due to heat of hydration, starting from the initiation of hydration and continuing for 5 days. A total of 101 time steps were recorded per sample, therefore allowing for a resolution of 0.05 days.

A wide range of parameters is varied per sample, with the aim of producing a robust TIP prediction model capable of predicting defects in piles of arbitrary characteristics, both inside and outside the reinforcement cage. The set of geometric, thermal and hydration-related parameters varied across the simulations is summarised in Table 1. All material properties are assumed to be temperature-independent.

Each parameter combination was generated using a Latin hypercube sampling (LHS) technique, a stratified sampling approach that ensures a well-distributed representation of the input space. LHS allows for efficient exploration of high-dimensional parameter spaces by ensuring that each sample is equidistantly placed along each input dimension, reducing redundancy and improving the parametric coverage across all sample instances (McKay *et al.* 1979).

Defect sizes were randomly assigned by sampling the radius ratio (defect radius relative to pile radius) from a uniform distribution between 0.1414 and 0.8. This corresponds to a defect cross-sectional area ranging from 2% to 64% of the pile cross-section at its widest point. As the randomly sampled defect radii are subsequently raised to the power of three when computing spherical volumes, the distribution naturally favours the generation of defects with smaller volumes. This aligns with defect sizes expected from TIP studies, where defects as small as 8% of a pile cross-sectional area are detectable (Ashlock and Fotouhi 2014), and large defects corresponding to 18.5% of the pile cross-sectional area have been studied (Mahjoubi *et al.* 2025). The average defect volume across all samples is 0.315 m³, with the median volume being 0.104 m³, corresponding to a sphere with radius approximately 29 cm. As previously mentioned, in cases where the defect extends beyond the pile boundaries and into the surrounding soil, the defect was trimmed, and only the part located inside the pile is considered as defect, which naturally reduces the defect size. In this case, random sampling integration is used as a generic approximation of the defect volume and centroid,

Table 1. Numerical model parameters.

Parameter	Lower bound	Upper bound	Unit	Description
$\theta_{defect1,2,3}$	0.0	2π	[rad]	Angular position of defects 1, 2 or 3
$r_{d,def1,2,3}$	0.0	1.0	[-]	Ratio of radial distance of defects 1, 2 or 3 centroid to the pile centre.
$r_{def1,2,3}$	0.1414	0.8	[-]	Ratio of defects 1, 2 or 3 radius to pile radius
$h_{def1,2,3}$	0.15	0.8	[-]	Ratio of depth centroid within layers 1, 2 and 3
$\lambda_{soil1,2,3}$	0.5	2.0	[W/(m K)]	Thermal conductivity of soil 1, 2 and 3 (Farouki 1981, Wu <i>et al.</i> 2025)
λ_{conc}	1.0	2.5	[W/(m K)]	Thermal conductivity of concrete (Asadi <i>et al.</i> 2018)
$C_{p,soil1,2,3}$	1200	2400	[kJ/(m ³ K)]	Volumetric heat capacity of soil (Kodesová <i>et al.</i> 2013)
$C_{p,conc}$	2000	2500	[kJ/(m ³ K)]	Volumetric heat capacity of concrete (Tatro 2006)
r_{pile}	0.3	1.2	[m]	Pile radius
L_{pile}	20	40	[m]	Pile length
L_1/L	0.15	0.45	[-]	Ratio of thickness of soil layer 1 to pile length
L_2/L	0.15	0.45	[-]	Ratio of thickness of soil layer 2 to pile length
θ_{cage}	0.0	2π	[rad]	Angular position of the cage shift
r_{cage}	0.0	0.065	[m]	Radial distance of cage shift
β	0.525	0.875	[-]	HoH model fitting parameter (Liu 2022a)
τ	52500	87500	[s]	HoH model fitting parameter (Liu 2022a)
Q_{max}	225	375	[kJ/kg]	Maximum heat generation capacity of concrete mix (Liu 2022a)
T_{pour}	10	25	[C°]	Initial (concrete pour) model temperature

evaluating five million points to ensure accuracy (Lemieux 2009). Material properties are then assigned accordingly to the pile and soil domains.

In order to account for potential cage misalignments arising from potential poor construction, two parameters θ_{cage} and r_{cage} are modelled to factor in the polar angle and radial displacement of the cage. These affect probe placement and the resulting thermal evaluation.

Following the guidance from ASTM (2025), eight probe columns were placed radially around the pile centre at equidistant angles simulating the attachment onto cage rebars, allowing for 75 mm of concrete cover. While ASTM recommendations typically suggest a longitudinal probe spacing of approximately 300 mm, computational constraints required an increased spacing of between 400 and 800 mm, depending on the pile length. A fixed depth resolution of 50 probes was used for all simulations to allow for consistent data generation irrespective of the pile length, totalling up to 400 probes per evaluated pile. This resolution is clearly conservative, as reduced spacing (either following the 150–500 mm recommended spacing in ASTM (2025) or the near-continuous measurements provided by fibre optics) increases the available data and therefore generally facilitates the detection of defects.

In order to further improve the robustness of the database, potential instrument misreadings were accounted for by adding an artificial, normally distributed stochastic noise with a standard deviation of 0.1°C to all temperature readings prior to ANN training. This value was selected to reflect the typical accuracy range reported for temperature sensing technologies used in pile monitoring applications (Sánchez Fernández *et al.* 2025). This regularisation step allows better generalisation and reduced overfitting in the two deep learning models.

2.3. Machine learning model

Two deep learning models were developed in this study to analyse and interpret temperature data obtained from TIP. The first model is a multi-label classifier designed to predict the presence or absence of defects within a concrete pile. It performs three parallel binary classifications, each corresponding to one of the possible defect zones along the shaft. The second model is a regression framework that estimates detailed defect characteristics, including defect location and size, reinforcement cage displacements and hydration-related parameters, by making use of the full spatio-temporal temperature field, probe coordinates and geometric properties of the pile.

Both models were implemented using the TensorFlow and Keras libraries in Python (Chollet 2021). Their final architectures were selected after extensive experimentation involving variations in layer types, depth, filter dimensions, kernel sizes, dilation rates, learning rates, dropout values and data preprocessing strategies.

2.3.1. Data processing and ML inputs

Both the classifier and regressor models make use of the full sample database, using an 80–10–10 split for model training, validation and testing. These use the spatio-temporal data obtained from the FE model stage. This produces a tensor with dimensions (N,T,50,8,C), corresponding to the number of samples (N), time steps (T), temperature sensors per depth (50), probe alignments (8) and data channels (C), respectively. The values of T and C vary depending on the specific architecture used in the classifier or regressor and will be discussed in detail in their respective sections.

In addition to the temperature field data, both models incorporate five static inputs per sample, namely the pile radius (r_{pile}) and length (L_{pile}), temperature of the pour (T_{pour}) and soil stratigraphy in the form of nominal values of the two randomised layer thicknesses (L_1 and L_2). These inputs capture the essential geometric and thermal properties of the samples.

The regression model includes an additional input: a spatial mapping of Cartesian coordinates for each temperature sensor. This coordinate grid has dimensions (50, 8, 3), corresponding to the x , y , and z positions of all 400 sensors arranged along the pile shaft. This spatial mapping provides positional context for the network's defect location predictions.

2.3.2. Classifier

The proposed classifier processes temperature data spanning 61 time steps, corresponding to a monitoring period of 0.1 to 3.1 days after concrete pouring, with measurements recorded every 0.05 days. To enhance

the model's sensitivity to the presence or absence of defects, two additional data channels were incorporated into the input: the mean and maximum temperatures computed across the hydration period. This results in a final input shape of (61, 50, 8, 3), corresponding to time steps, sensor depths, sensor alignments and channels, respectively. Prior to model training, both the temperature data and the five geometric and geotechnical characteristics were standardised using z-score normalisation.

The architecture consists of five TimeDistributed two-dimensional convolutional layers, each with a kernel size of 3×3 , 'same' padding and ReLU activations. Dilation rates were varied as [1,2,4,2,1], and the number of filters increased from 32 to 128 across the layers. Batch normalisation was applied after each convolutional block to stabilise and accelerate the training process.

To capture the spatio-temporal dependencies in the data, the output of the convolutional blocks was passed to a two-dimensional ConvLSTM layer with 64 filters. This was followed by a TimeDistributed global max pooling layer to retain the most salient spatial features at each time step. A subsequent LSTM layer with 128 units and a 30% dropout rate summarised the temporal features.

The static input branch consists of two fully connected layers with 50 and 30 units, respectively, both using ReLU activations and a 20% dropout rate between layers. The outputs of the temporal and static branches were concatenated and passed through an additional fully connected layer with 100 units, followed by a 40% dropout layer for regularisation. Finally, a sigmoid-activated output layer with three neurons was used to predict the presence or absence of defects in each of the three soil layers.

To ensure robust performance, 10-fold cross-validation was implemented. Within each fold, 10% of the entire dataset was held out as the validation set. A batch size of 30 samples was used at the training stage. Early stopping with a patience of 15 epochs was applied to prevent overfitting, alongside a learning rate reduction scheme (factor of 0.75, patience of 8 epochs, minimum learning rate of $1e-7$) to promote convergence. A complete summary of the model architecture is illustrated in Figure 2.

2.3.3. Regressor

In contrast to the classification model, the regression network is designed to predict continuous-valued defect characteristics and structural parameters for each sample. Specifically, it estimates the 3D cartesian coordinates and volumes of up to three potential defects, along with the displacement of the reinforcement cage (x_{cage}, y_{cage}), obtained from the input polar displacements (θ_{cage}, r_{cage}), and three concrete hydration parameters (β, τ, Q_{max}), resulting in a total of 17 regression targets.

Due to the increased complexity and the wider temporal dataset required for accurate defect localisation and characterisation, the full 5 days following concrete pouring are used as input,

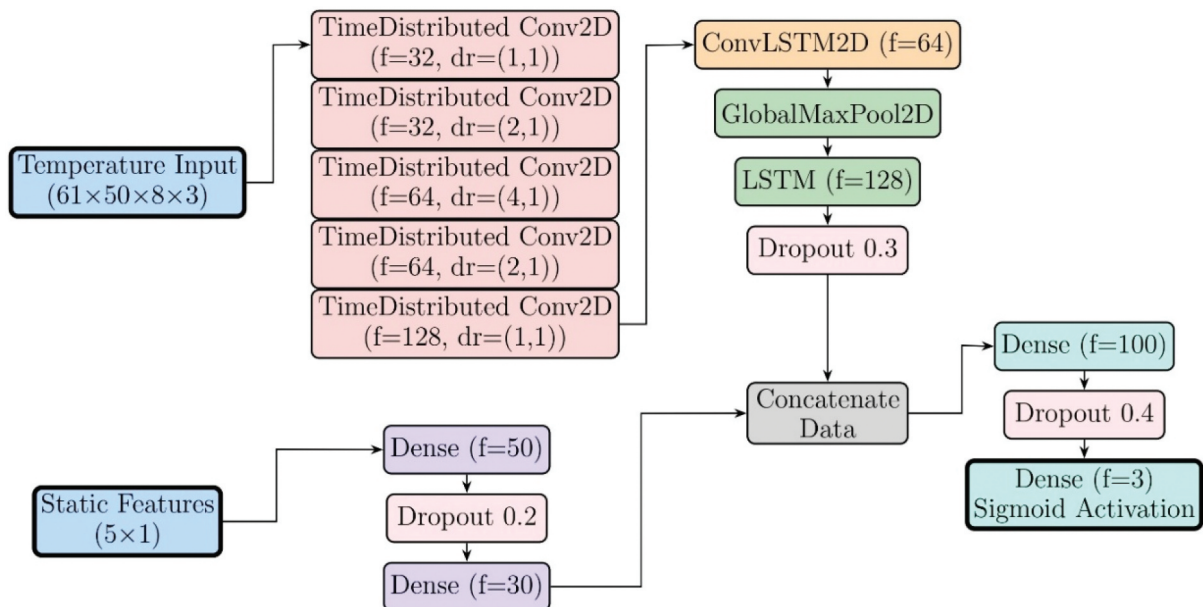


Figure 2. ANN architecture flowchart for the classifier.

sampled every 0.1 days. This results in a temporal resolution of 51 time steps per sample. A single channel representing the temperature increase over time is provided to the model, forming a spatio-temporal input of shape (51,50,8,1) per sample. The regressor also incorporates two additional input branches: the spatial coordinates of each probe (a tensor of size 50,8,3) and the five static parameters per sample (5 parameters).

The architecture of the model consists of three parallel input branches:

- (1) A temperature branch, where the temperature input is passed through a sequence of eight TimeDistributed convolutional blocks. Each block uses 2D convolutions with varying dilation rates, ranging between 1 and 3 in the depths, and between 1 and 2 in the alignments, therefore enabling the network to capture localised patterns at multiple spatial scales. The number of filters increases from 80 to 400. Batch normalisation is applied after each convolutional layer to improve stability. This is followed by a TimeDistributed global average pooling layer, which compresses the spatial dimensions per time step. The resulting time series of feature vectors is then processed by a single LSTM layer with 250 units, enabling the network to learn temporal dependencies across the hydration period.
- (2) A coordinate branch, where the 3D spatial coordinates of each probe are processed using two 2D convolutional layers with dilation rates (2,1) and 100 and 150 filters, respectively. These layers are followed by batch normalisation and a global average pooling layer, resulting in a fixed-length feature vector that encodes spatial information about the probe configuration for each of the samples.
- (3) A static feature branch, where the five input variables are passed through two fully connected layers with 120 and 60 units, respectively, both using ReLU activation functions.

The outputs of the three branches are concatenated and passed through a regression head composed of two dense layers with 400 and 200 units, respectively, separated by a dropout layer with a rate of 40%. A final dense layer with linear activation outputs the 17 continuous predictions for each sample.

All inputs are standardised using StandardScaler, based on the training set of each fold. The network is trained using a mean squared error (MSE) function. The training is conducted using the Adam optimiser with an initial learning rate of 1e-3, a batch size of 40 and a maximum of 200 epochs. Early stopping with a patience of 15 epochs and a learning rate scheduler (reduction factor of 0.66, patience of 8 epochs, minimum learning rate of 1e-7) are employed to ensure convergence. Similar to the classifier, a 10-fold cross-validation was also implemented for the regressor. A complete schematic of the regression model architecture is presented in [Figure 3](#).

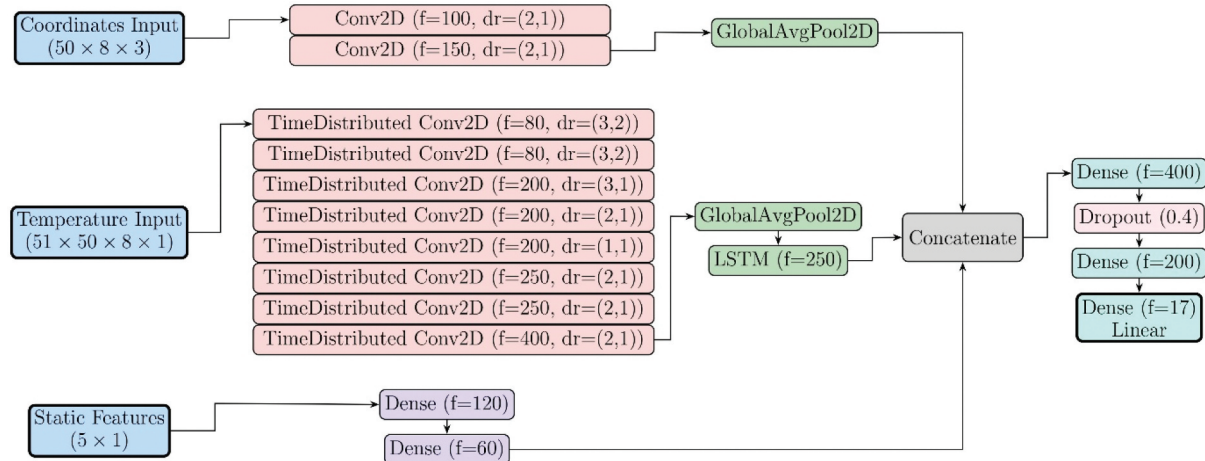


Figure 3. ANN flowchart for the regressor.

3. Results & discussion

3.1. Classification ANN

A total of 1,944 samples were used to evaluate the performance of the proposed multi-defect classification ANN. The classifier outputs three binary predictions per sample, and Figure 4 therefore presents the confusion matrices for each of the three defect labels. High classification performance is observed across all three defects per sample, with test-set accuracies reaching 94.3% for Defect 1, 92.5% for Defect 2 and 93.8% for Defect 3. This results in an average accuracy of 93.6% for the test set for this particular fold, a level similar to that reported by Sánchez Fernández *et al.* (2025) for the simpler 2D TIP analysis. Other CNN concrete defect classifiers achieve a 94.4% accuracy on bulging, weak concrete and neck defects (Wu *et al.* 2022).

Most of the mislabelled samples correspond to false negatives. In the training set, an average of 95.5% samples are correctly labelled, therefore indicating negligible overfitting. Considering the overall model performance across all folds, an average of $92.8\% \pm 1.0\%$ standard deviation is obtained across the 10 folds. When evaluated on a per-sample basis, where a prediction is considered correct if at least one defect (or the absence of all defects) is correctly identified, the model achieves a higher accuracy of 96.8%, compared to the stricter per-defect metric.

To further examine the behaviour of the classifier across prediction confidence thresholds, Figure 5 illustrates how the number of false positives (FP), false negatives (FN) and total misclassifications vary as the threshold is adjusted. Although a threshold of 0.5 yields the lowest combined error count and is therefore adopted throughout this paper, the threshold can be deliberately modified, on the trained classifier, depending on the engineering design objectives. In safety-critical civil engineering applications,

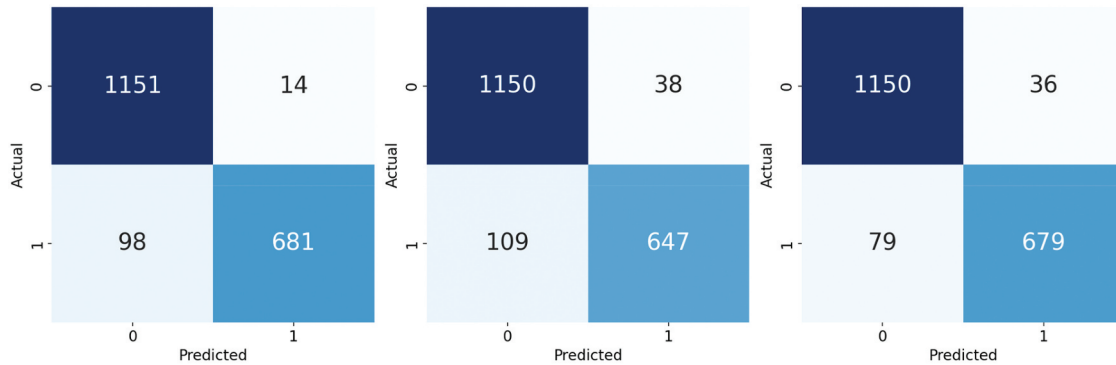


Figure 4. Confusion matrices for defects 1, 2 and 3.

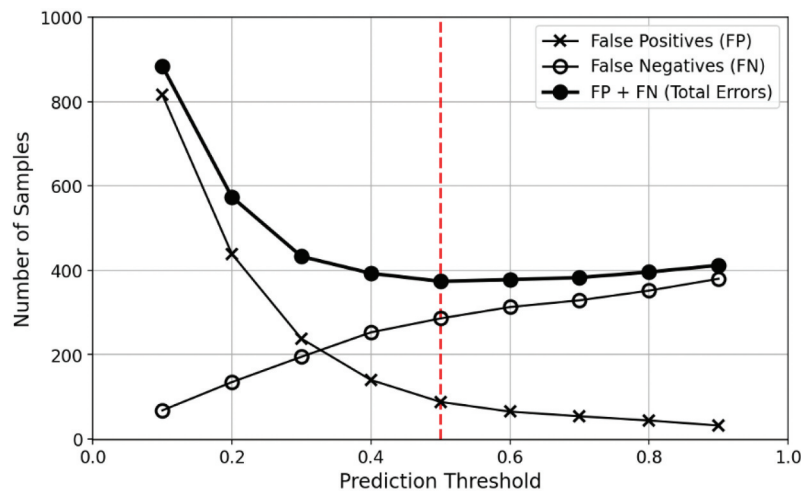


Figure 5. Number of false positives and negatives as a function of the prediction threshold.

a conservative detection strategy may be preferred, where the risk of missing a defect (FN) must be minimised. This can be achieved by lowering the threshold, thereby classifying more samples as defective. Conversely, for projects where unnecessary interventions carry significant cost or programme implications, a less conservative strategy can be adopted by increasing the threshold and reducing FP at the cost of allowing a higher proportion of FN.

Figure 6 provides a heatmap showing spatial distributions of the classifier's false-negative predictions in the pile cross-sectional plane. These results show that most false negatives tend to cluster around the central region of the pile cross-section, suggesting that central defects are more challenging to detect as they lead to subtler thermal profile deviations due to the larger distances to the circumferentially placed probes. In contrast, defects located towards the edge of the piles are predicted with higher accuracy. Despite this, the model is shown to be able to predict with overall high accuracy the existence of defects not only in regions outside of the reinforcement cage but also those defects in the central part of the pile, traditionally considered more challenging to detect (Boeckmann and Loehr 2019).

To explore this further, Figure 7 plots predicted classification probabilities against actual defect volumes for each defect label, showing the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The visualisation is limited to a defect volume range between 0 and 0.10 m^3 , to highlight the low-volume range where 97.6% of false negatives occur. Furthermore, false negatives are primarily concentrated in the lowest volume range, with the majority lying below 0.005 m^3 . These defects are thermally indistinguishable from healthy regions and, from an engineering perspective, could be considered non-critical. Supporting this, the average and median volume for false negatives are 0.015 m^3 and 0.006 m^3 , respectively. By contrast, true positives logically show substantially larger volumes and radii, with a median volume of 0.141 m^3 and median radius of 0.338 m . This distinction matters to prove the model's ability to detect structurally relevant defects while filtering out anomalies with minimal thermal, geometric or,

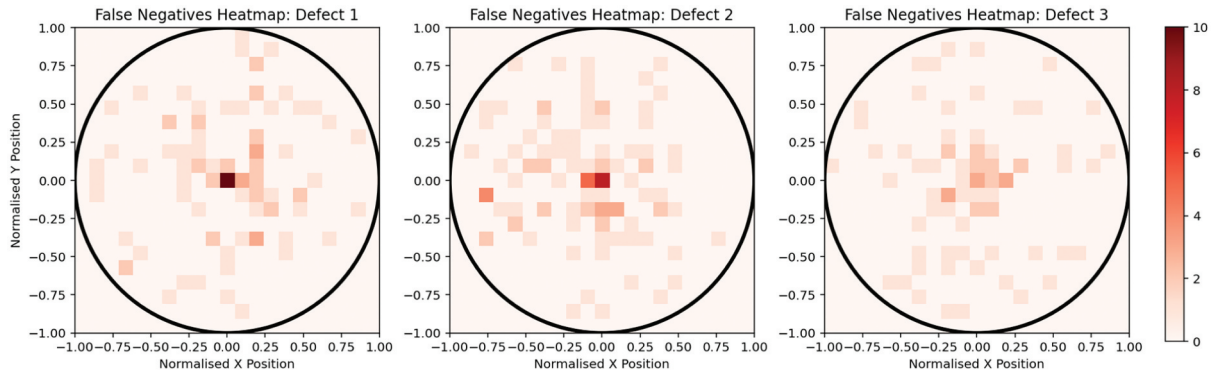


Figure 6. Heatmap of the X-Y false negative location for defects 1, 2 and 3.

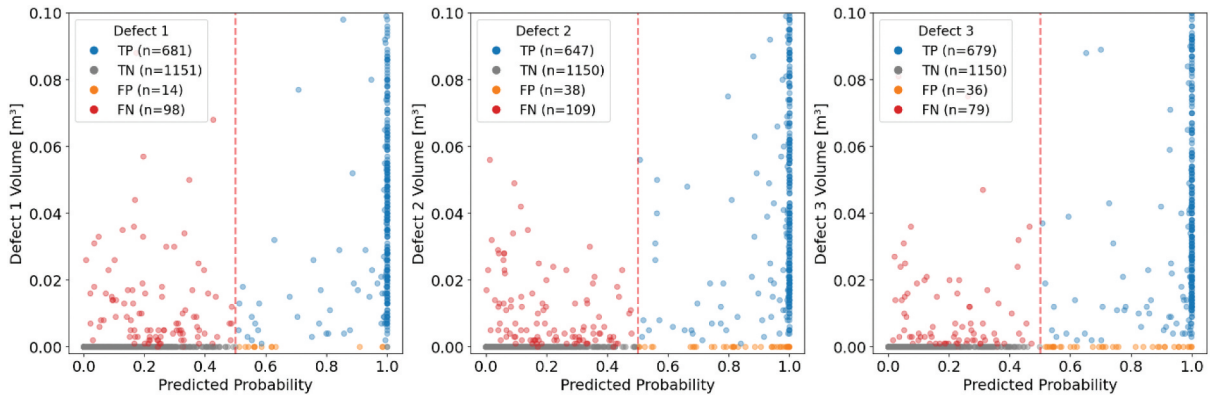


Figure 7. Classification output of small test set defects compared to the volume of defects 1, 2 and 3.

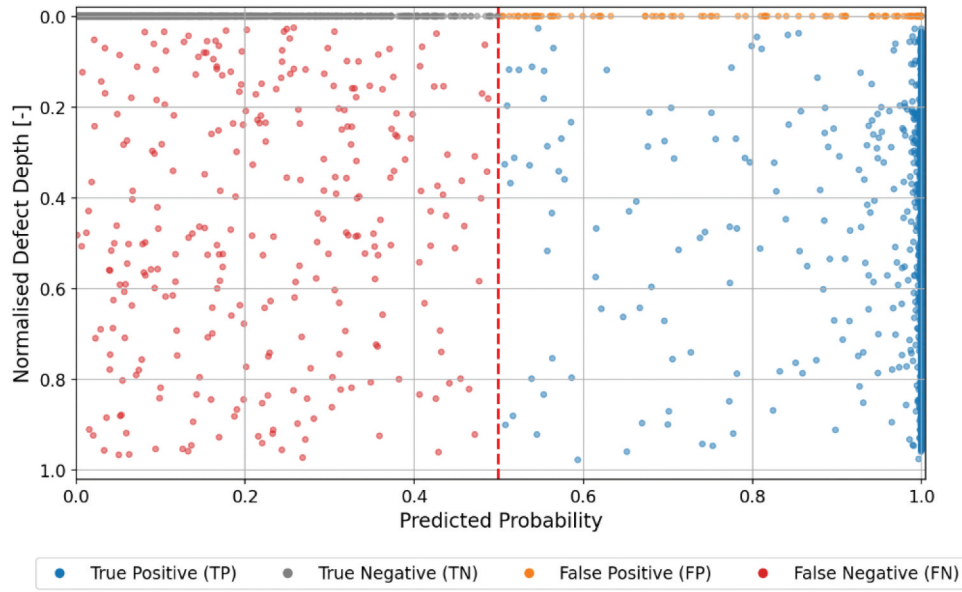


Figure 8. Predicted probability per defect against defect depth.

indeed, mechanical significance. For context, the average and median volume of the entire defect dataset are 0.315 m^3 and 0.104 m^3 , respectively, well in excess of the characteristic values of false-negative samples.

Figure 8 illustrates the relationship between the classifier's predicted probability and the normalised defect depth for all defects in the test set. Several trends emerge. First, the model produces high-confidence predictions (probabilities close to 0 or 1) for a large proportion of samples, with most true positives clustering near probability 1 and most true negatives near probabilities 0–0.3. This indicates that the model is not only correctly classifying the majority of cases but is generally confident in its decisions for defects across all depths. The distributions of false negatives and false positives are evenly distributed along the pile length, rather than being concentrated at any specific depth. This suggests that the model does not exhibit blind spots along the pile length. Instead, FN events are mostly associated with small or low-contrast defects, consistent with the earlier volume analysis which showed that FN typically correspond to very small physical anomalies.

To improve interpretability of the classifier model, a feature importance analysis was conducted using SHAP (SHapley Additive exPlanations), a game-theory-based method commonly employed for neural network interpretation (Lundberg and Lee 2017). SHAP values were computed using the GradientExplainer on the trained TensorFlow classifier, reconstructing the training fold and scaling procedure to ensure consistency.

Given the high dimensionality of the temperature input tensor (61 time steps $\times$ 50 probes $\times$ 8 columns $\times$ 3 channels), the SHAP values were aggregated using the mean absolute contribution, averaged across samples, time, probes and radial positions. Static inputs were similarly aggregated to obtain a mean SHAP value contribution per feature.

Figure 9 presents the mean absolute SHAP contribution per input element. The results indicate that, among the temperature inputs, elements from the raw temperature channel consistently provide the largest contribution to defect detection, followed by the derived mean and maximum temperature channels. Among the static inputs, geometric parameters, particularly the pile length and defect depth ratios, show stronger influence than the pouring temperature. This suggests that the classifier primarily relies on physically meaningful geometric descriptors in combination with temporal thermal evolution, supporting the consistency of the model's learned behaviour with expected dynamics.

In summary, the classification ANN demonstrates strong predictive capabilities across all defect depths, with high per-label and sample-level accuracies. Moreover, the classifier exhibits useful decision-making behaviour, where lower-confidence predictions are predominantly associated with very small or centrally located defects (i.e. cases that are less thermally distinguishable and less structurally critical).

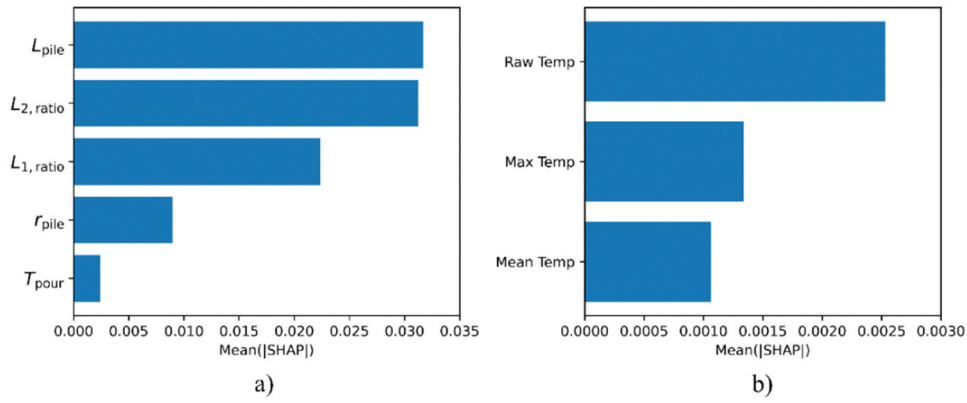


Figure 9. Mean SHAP feature importance for (a) static features and (b) temperature features.

Table 2. Regressor metrics for cross-validation fold.

Variable	MAE	RMSE	R2	R2 (No FN or FP)
β [–]	0.014	0.018	0.967	0.969
τ [s]	2755	3593	0.875	0.88
Q_{max} [$kJ \cdot kg^{-1}$]	15.3	18.9	0.816	0.82
Cage misalignment (x) [m]	0.002	0.002	0.992	0.993
Cage misalignment (y) [m]	0.002	0.002	0.992	0.992
Defect 1 Position (x) [m]	0.032	0.076	0.842	0.905
Defect 1 Position (y) [m]	0.030	0.069	0.841	0.91
Defect 1 Depth (z) [m]	0.491	1.184	0.811	0.931
Defect 1 Volume [m^3]	0.034	0.084	0.949	0.95
Defect 2 Position (x) [m]	0.034	0.078	0.809	0.881
Defect 2 Position (y) [m]	0.033	0.077	0.817	0.91
Defect 2 Depth (z) [m]	1.553	3.772	0.730	0.908
Defect 2 Volume [m^3]	0.03	0.08	0.936	0.933
Defect 3 Position (x) [m]	0.029	0.068	0.861	0.907
Defect 3 Position (y) [m]	0.028	0.067	0.846	0.906
Defect 3 Depth (z) [m]	2.389	5.634	0.792	0.92
Defect 3 Volume [m^3]	0.029	0.075	0.955	0.955
Average Performance	–	–	0.872	0.922

3.2. Regressor ANN

The regression model was trained using the same 10-fold cross-validation splits as the classification model. Table 2 presents the average mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination (R^2) for all 17 predicted outputs. In this fold, the model achieves an average R^2 of 0.872 for all variables. This result closely aligns with the overall cross-validated performance of 0.873 ± 0.004 across the 10 folds. These values can be interpreted as a measure of the model's uncertainty for unseen inputs. The average training accuracy for the model is 0.969 ± 0.018 , showing moderately low overfitting levels given the complexity of the problem at hand.

Model performance is particularly strong for defect volume predictions, where R^2 values average 0.947 and MAE remained below $0.035 m^3$, showing excellent prediction accuracy. Similarly, predictions for the misalignment of the reinforcement cage show robust performance, with R^2 values above 0.99 and MAE of merely 2 mm. Regarding the concrete hydration parameters, these can be predicted with high accuracy, particularly the parameter determining the start of hydration process, β .

The predictions of the Cartesian coordinates defining the location of the centre of the defect, particularly in the z direction, exhibited lower accuracy compared to other outputs. This reduction in performance is primarily attributed to the same subset of samples which had proven difficult to be correctly labelled by the classifier. These particularly ambiguous cases also negatively impact the overall performance of the regression model, since samples deemed to be non-defective are assigned centroid coordinates of zero. Therefore, when these are mislabelled, the mismatch leads to large errors, particularly in the z-coordinate output. To establish the model's effective predictive capacity, regression metrics were recalculated excluding all samples for which the classifier had provided either a false negative or a false positive. Under these filtered

conditions, the average R^2 for the z-coordinate outputs increased significantly to 0.920, demonstrating that the previously observed errors were predominantly due to the presence of intrinsically undetectable samples, rather than issues in the regression model itself. This is shown in an additional column of Table 2.

Figure 10 presents scatter plots of predicted versus true values for all 17 regression outputs on the test set, with individual samples coloured according to the classification outcome: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). When looking at the z-coordinate predictions for the three

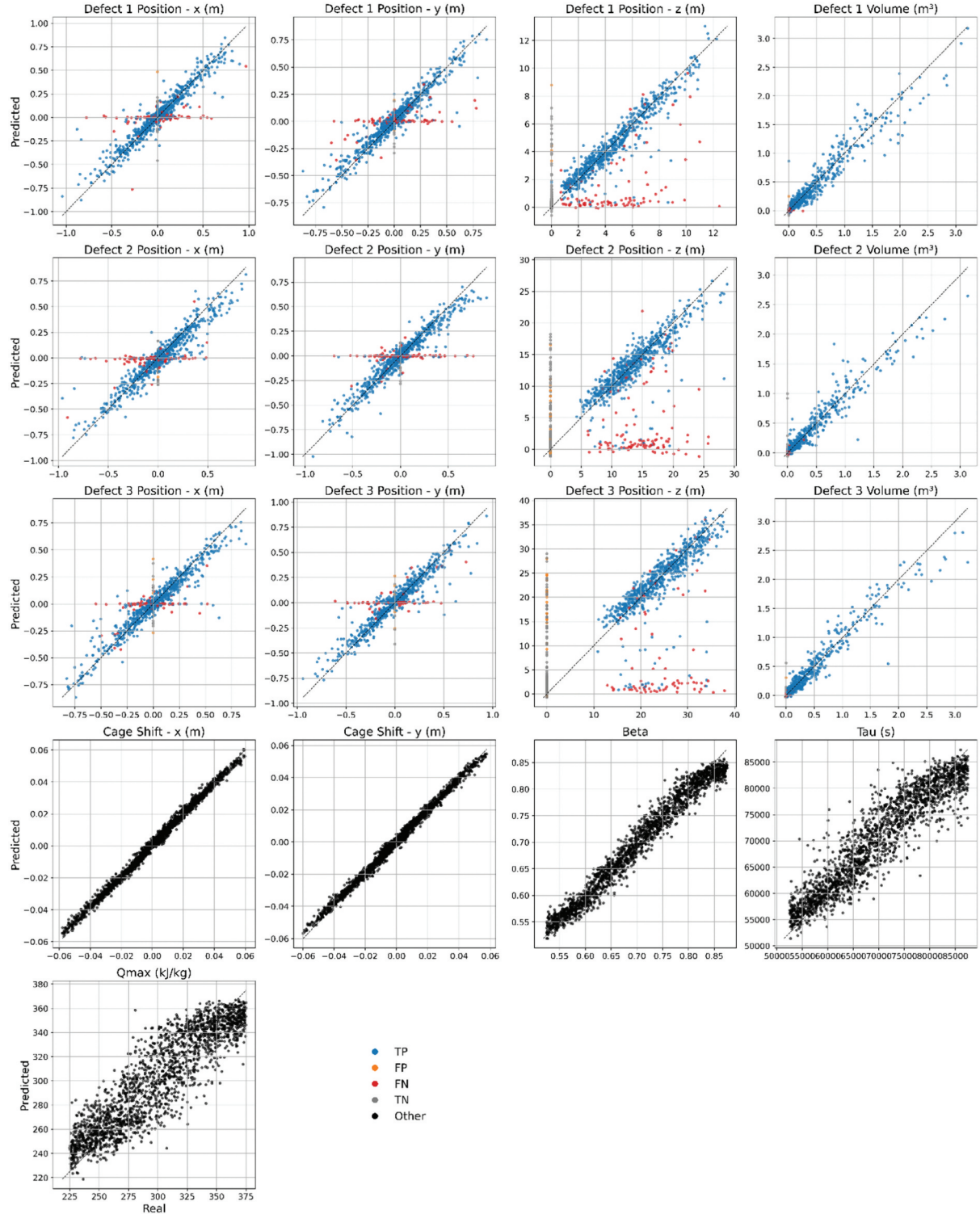


Figure 10. Real against predicted values for all outputs of the test set.

defects, the red-labelled false negatives cluster around the zero predicted area and are also miscalculated by the regressor, increasing the calculated average error. In contrast, x and y coordinate predictions are less affected, as even when classification errors occur, the true values of these quantities are limited to the interval $[-1.2, 1.2]$, reducing the impact on error metrics. The defect volumes and reinforcement cage displacements are accurately calculated for all cases, with tight clustering around the 1:1 line. Similarly, the three hydration parameters are estimated accurately, with trends being captured by the model, even if greater scatter is observed for τ and Q_{max} . This behaviour is consistent with the underlying formulation of the hydration model, in which different combinations of τ and Q_{max} can produce similar temperature responses at the sensors within the evaluated timeframes (0–5 days). Since all temperature measurements are taken along radial lines, the inverse problem is effectively constrained by a single set of temperature readings so that multiple combinations remain indistinguishable. As a result, τ and Q_{max} are intrinsically harder to identify uniquely, even when the overall response is reproduced well by the surrogate model.

4. Conclusions & future work

This study presents a novel ML framework comprising classification and regression of deep neural networks designed to automatically interpret three-dimensional TIP data for the detection and characterisation of defects in concrete pile foundations. Both models were trained on a numerically generated dataset covering a broad and fully randomised range of geometrical, material and thermal parameters that reflect realistically the variability encountered in construction and ground conditions. The following conclusions can be drawn:

- The classifier achieved an average defect detection accuracy of 93.6% across all test folds, demonstrating its effectiveness in identifying the presence or absence of up to three defects per pile.
- Most misclassified defects were located in the central regions of the pile, which also correspond to areas furthest from the temperature probes. This highlights the challenge of detecting defects that are spatially distant from TIP sensor locations, where temperature variations are either too small to be measured and are therefore masked by the stochastic noise, or cannot be detected at all for the duration of the hydration period.
- The model exhibited a higher false-positive than false-negative rate using the standard classification threshold, although this can be tuned depending on the user's operational priorities. For example, the model allows users more concerned with avoiding missed defects to lower the classification threshold, accepting more false positives in return for improved sensitivity.
- The regression model demonstrated near-perfect accuracy in estimating the reinforcement cage displacement, with an average R^2 of 0.992 for both x and y components. Similarly, the volume of the defects and the hydration parameters (β , τ , Q_{max}) were predicted with high precision.
- The model initially achieved an R^2 of approximately 0.82 for the 3D cartesian coordinates of the centre of the defects. However, this metric improved significantly to an average R^2 of 0.91 after excluding samples that the classifier had also failed to classify correctly. This indicates that the regression model is highly robust and that a large proportion of the coordinate prediction error arises from a small subset of cases that are fundamentally difficult to detect using TIP data, as these are defects of 0.005 m^3 in volume on average.

The objective of this paper is to present a proof-of-concept framework, which resulted in two 'general models' for identifying the presence and characteristics of defects in a pile. To ensure their applicability to a wide range of scenarios, several techniques were applied, including extensive parametric variations of material and geometric properties, multiple-defect scenarios, reinforcement cage displacement and stochastic noise in the temperature data. While these measures support generalisation, an important next step would entail a validation against field results. Currently, a limited number of tests have been reported in the literature (e.g. Rui *et al.* (2017), Sun *et al.* (2021)), with none providing sufficiently detailed datasets that enable the direct use of the developed 'general models'.

In summary, the proposed ML-based approach offers a fast, automated and accurate method for interpreting TIP data, significantly reducing reliance on subjective manual analysis. The classifier

proves to be a reliable pre-screening tool for identifying the presence or absence of defects at any depth along the pile, enabling rapid and objective assessment of structural integrity. When combined with a robust regression model capable of accurately estimating defect size, three-dimensional location and reinforcement cage misalignment, the result is a comprehensive and reliable workflow for defect detection in concrete piles. The flexibility of the proposed ANNs, being able to adapt to different construction scenarios, and strong predictive capability for both classification and regression tasks make it a promising tool for advancing quality control and defect characterisation in foundation engineering.

Author contributions

CRedit: **Javier Sánchez Fernández**: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft; **Agustín Ruiz López**: Conceptualization, Project administration, Supervision, Writing – review & editing; **David M. G. Taborda**: Conceptualization, Project administration, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under grants [EP/V042149/1] (SaFEGround – Sustainable, Flexible and Efficient Ground-source Heating and Cooling Systems) and [EP/X52556X/1] (SIDeTools – Sustainable Infrastructure Design Tools), and by the Natural Environment Research Council (NERC) under grant [NE/S007415/1].

Data availability statement

The data that support the findings of this study are available in Zenodo at <https://doi.org/10.5281/zenodo.17897881> (Sánchez Fernández 2025b).

References

- Asadi, I., *et al.*, 2018. Thermal conductivity of concrete – a review. *Journal of Building Engineering*, 20 (July), 81–93. doi:10.1016/j.jobe.2018.07.002
- Ashlock, J.C. and Fotouhi, M.K., 2014. Thermal integrity profiling and crosshole sonic logging of drilled shafts with artificial defects. In: M. Abu-Farsakh, X. Yu, and L. Hoyos, *Geo-Congress 2014: Geo-characterization and Modeling for Sustainability*, Atlanta, Georgia. American Society of Civil Engineers, 1795–1805. doi:10.1061/9780784413272.176
- ASTM, 2016a. D5882 - Standard test method for low strain impact integrity testing of deep foundations. *Annual Book of ASTM Standards*. doi:10.1520/D5882-16
- ASTM, 2016b. D6760 - Standard test method for integrity testing of concrete deep foundations by ultrasonic crosshole testing. *American Society for Testing and Materials*. doi:10.1520/D6760-16.1
- ASTM, 2025. D7949-25 - Standard test method for integrity testing of concrete deep foundations by thermal methods. *American Society for Testing and Materials* (April), 1–7. doi:10.1520/D7949
- Boeckmann, A.Z. and Loehr, J.E., 2019. Evaluation of thermal integrity profiling and crosshole sonic logging for drilled shafts with concrete defects. *Transportation Research Record: Journal of the Transportation Research Board*, 2673 (8), 86–98. doi:10.1177/0361198119842112
- Chollet, F., 2021. *Deep learning with python*. New York, United States: Manning Publications.
- Farouki, O.T., 1981. *Thermal properties of soils*. Clausthal-Zellerfeld, Germany: Trans Tech Publications.
- Finno, R.J., *et al.*, 2002. Non-destructive evaluation of drilled shafts at the Amherst NGES test section. In: M.W. O'Neill and F.C. Townsend, eds. *Deep foundations 2002*. Reston, VA: American Society of Civil Engineers, 883–897. doi:10.1061/40601(256)62. [Accessed 26 February 2002].
- Klingmüller, O. and Kirsch, F., 2004. A quality and safety issue for cast-in-place piles - 25 years of experience with low-strain integrity testing in Germany: from scientific peculiarity to day-to-day practice. In: J.A. DiMaggio and M. H. Hussein, eds. *Geo-Trans 2004*, Los Angeles, CA, 27–30 July 2004. Reston, VA: American Society of Civil Engineers, 202–221.

- Kodesová, R., *et al.*, 2013. Thermal properties of representative soils of the Czech Republic. *Soil and Water Research*, 8 (4), 141–150. doi:10.17221/33/2013-swr
- Lemieux, C., 2009. Monte Carlo and quasi-Monte Carlo sampling. In: *Springer series in statistics*. 1st ed. New York, NY: Springer New York, 369. doi:10.1007/978-0-387-78165-5
- Liu, R., 2022a. *Numerical modelling of complex thermo-hydro-mechanical interactions in thermo-active pile foundations*. Thesis (PhD). Imperial College London.
- Liu, R., *et al.*, 2022. A practical heat of hydration model for concrete curing for geotechnical applications. *Geotechnical Research*, 9 (1), 23–31. doi:10.1680/jgere.21.00054
- Loseva, E., Lozovsky, I., and Zhostkov, R., 2022. Identifying small defects in cast-in-place piles using low strain integrity testing. *Indian Geotechnical Journal*, 52 (2), 270–279. doi:10.1007/s40098-021-00583-y
- Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30 (2), 4766–4775.
- Mahjoubi, S., Sun, M., and Lin, C., 2025. Thermal integrity profiling of cast-in-place piles with defects: insights from numerical modeling. *Journal of Geotechnical and Geoenvironmental Engineering*, 151 (9), 1–19. doi:10.1061/jggefk.gteng-13631
- McKay, M.D., Beckman, R.J., and Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42 (1), 55–61. doi:10.1080/00401706.2000.10485979
- Monteiro, P.J.M., Miller, S.A., and Horvath, A., 2017. Towards sustainable concrete. *Nature Materials*, 16 (7), 698–699. doi:10.1038/nmat4930
- Mullins, G., 2010. Thermal integrity profiling of drilled shafts. *DFI Journal - The Journal of the Deep Foundations Institute*, 4 (2), 54–64. doi:10.1179/dfi.2010.010
- Mullins, G., *et al.*, 2007 May. *Thermal integrity testing of drilled shafts*. Tallahassee, FL: Florida Department of Transportation.
- Mullins, G. and Winters, D., 2011. Infrared thermal integrity testing quality assurance test method to detect drilled shaft defects. 2011. Available from: <https://api.semanticscholar.org/CorpusID:106876614>
- Parrot, L. and Killoh, D., 1984. Prediction of cement hydration. *British Ceramic Proceedings*, 35, 41–53.
- Pratt, Q., 2016. *Material quantities of foundation systems in building structures*. Cambridge, MA: MIT.
- Qiu, P., *et al.*, 2025. Pile integrity testing using non-destructive testing techniques and artificial intelligence: a review. *Applied Sciences*, 15 (15), 8580. doi:10.3390/app15158580
- Rui, Y., *et al.*, 2017. Integrity testing of pile cover using distributed fibre optic sensing. *Sensors (Switzerland)*, 17 (12), 2949. doi:10.3390/s17122949
- Sánchez Fernández, J., 2025a. *Thermal integrity profiling modelled data for ML training* [Dataset]. Geneva, Switzerland: Zenodo. doi:10.5281/zenodo.14674381
- Sánchez Fernández, J., 2025b. *Three-dimensional thermal integrity profiling numerical data for ML model training* [Dataset]. Geneva, Switzerland: Zenodo. doi:10.5281/zenodo.17897881
- Sánchez Fernández, J., Ruiz López, A., and Taborda, D.M.G., 2025. A novel machine learning-based approach to thermal integrity profiling of concrete pile foundations. *Data-Centric Engineering*, 6, e33. doi:10.1017/dce.2025.10013
- Schindler, A. and Folliard, K., 2005. Heat of hydration models for cementitious materials. *ACI Materials Journal*, 102 (1), 24–33. doi:10.14359/14246
- Sun, Q., *et al.*, 2021. Thermal integrity testing of cast in situ piles: an alternative interpretation approach. *Structural Health Monitoring*, 20 (5), 2493–2512. doi:10.1177/1475921720960042
- Swaddiwudhipong, S., Chen, D., and Zhang, M.H., 2002. Simulation of the exothermic hydration process of Portland cement. *Advances in Cement Research*, 14 (2), 61–69. doi:10.1680/adcr.2002.14.2.61
- Tatro, S.B., 2006. Chapter 22: Thermal properties. In: J.H. Pielert, ed. *Significance of tests and properties of concrete and concrete-making materials*. West Conshohocken, PA: ASTM International, 226–237. doi:10.1520/STP37740S
- Wu, C.-S., *et al.*, 2022. Defect identification of concrete piles based on numerical simulation and convolutional neural network. *Buildings (Basel)*, 12 (5), 664–. doi:10.3390/buildings12050664
- Wu, S., *et al.*, 2024. Global CO₂ uptake by cement materials accounts 1930 - 2023. *Scientific Data*, 11 (1), 1–11. doi:10.1038/s41597-024-04234-8
- Wu, Y.H., *et al.*, 2025. Thermal conductivity of soil: a review on the vast experimental data and predictive models. *The International Journal of Thermal Sciences*, 208 (July 2024), 109486. doi:10.1016/j.ijthermalsci.2024.109486